Anonymous Referee #1

Thank you for your effort and valuable comments on our paper. Our responses are embedded below in blue.

This manuscript presents a quantification of CO emissions over Madrid, based on MOPITT measurements and WRF simulations. In my opinion, this paper represents an interesting work and it is a good complementary work of the study done by Pommier et al. (2013). I was very interested to read this paper especially by thinking that it is a good idea to use a model to optimize the estimation of the emissions. This part was lacking in the work previously done by Pommier et al. (2013). This paper fits perfectly for this journal. Nevertheless, the manuscript is not well structured, making sometimes difficult to read. Thus I recommend publication in ACP after the comments below are addressed.

Structure: Section 2.3.5: I do not understand why you present Fig A6 before A3, A4…
We chose to not refer to Fig A6 in section 2.3.5 since there is some information needed to understand this scatterplot correctly. The information is given in Appendix B, so this is where we introduce Fig A6, at the end of the document. We also moved some of the figures from the appendix to the main text, so there are no conflicts in the order of the presentation in our revised version.

The problem of organization is also shown with the caption in Fig A1. At this stage I do not know what the correction factor is. This factor is only mentioned from page 9. Moreover, in Fig. A1 the caption and the colors of the curves do not match. There is no dotted line. *Thank you for pointing this out*. *We changed the caption in the revised version and also changed the figure it now includes only the original WRF data without correction factor.*

It is odd to finish the paper by the sensitivity tests. These tests should be done before to analyze the results of the WRF optimization method. *We changed the order of presenting our paper in the revised version. The sensitivity tests are following the description of the WRF optimization method and are described in an extra section: 3.3.2.*

Page 10: Table A2 is described before Tab. A1. *we corrected the sequence.*
Difference with P13: The authors concluded – quoting the text: "the emission proxies in P13 are too optimistic". In the same time, they wrote that the RD can change up to 25% due to the mis-location of the city center. With a quick check with the work done by Pommier et al. (2013), we can see that the locations of the city used in this work do not match perfectly with the coordinates used in P13. For example, Sao Paulo is 23.54S, 46.64W in your work and 23.53S, 46.62W in P13.
Thank you for catching that mistake. The wrong coordinates were still in the tables. We updated the coordinates in our calculations to match them with Pommier, but forgot to update the coordinates listed in the tables. We updated the coordinates in the revised version.

This represents only a difference of 2 km but it seems even a difference of 0.7 km has an impact on the RD. It is interesting to see that P13 did not take into account this problem of location. It is probably a missing source of error in their study. Thus I agree with the authors the uncertainties in P13 are probably underestimated. Another remark about the differences between both studies: the differences may be explained by 3 parameters: - The resolution of the wind are not similar (0.75 in P13 vs 1deg in this work) - The PBLH (750 hPa in P13 vs 700 hPa in this study) – The filter used for the MOPITT data (cloud fraction = 0 and cloud index = 2 in P13 vs cloud diagnostic = 1, 2 in this study). How do use the pixels where there is a conflict between sea surface and land? P13 filtered out these data. The discrepancy

between both studies may decrease if similar criteria are used. *Thank you for considering these sources of differences. We have addressed these issues in some more detail in the revised version. We agree that the discrepancy might decrease if the exact same criteria were applied. The point we want to make in this section is that the method is very sensitive to slight differences in the filtered data. We did some extra tests to find out the importance of the PBLH and cloud fraction which we included in a new section: "Other sources of uncertainties". We used all the pixels which were according to the MOPITT filter land data, but we agree there could be a problem at the boundary of sea surface and land. We added the following sentence in the section "other sources of uncertainties":*

We do not filter MOPITT data for retrievals containing water bodies other than rejecting water and mixed retrievals using the standard MOPITT flags. Since MOPITT is not able to measure CO in the near-infrared over areas with low albedo, such as water, this can lead to biases in the emission trend estimates in our method. For Los Angeles and Sao Paulo, which are both close to the coast, our analysis may include some scenes with fractional areas of water, while P13 filtered these out. This might explain part of the difference in RD estimation seen in Fig. 5, especially for Sao Paulo.

Other major comments: Introduction: Is there any publications about the CO trend/pollution over Madrid? It will be informative to have a comparison of your results with previous studies.
Unfortunately we could not find any study on CO trend or other pollution over Madrid

Page 3, line 1: Pommier et al. 2013 did not quantify emissions. Estimate the change in the emissions is more appropriate. Clerbaux et al. 2008 did not calculate the emissions but they detected urban CO plumes. Thus delete this reference for this sentence. Then you can write, "Clerbaux et al. (2008) and Pommier et al. (2013) already demonstrated that…"
We changed the text according to your comments:
Furthermore, the first attempts have been made to use MOPITT CO retrievals to estimate emission changes over cities (Pommier et al., 2013). Clerbaux et al. (2008) and Pommier et al. (2013) demonstrated that CO pollution plumes over large cities can be distinguished from the background in satellite data.

Page 7, line 11: does it means you exclude the first days of your run? What is the period of your simulations? You should introduce this information before Section 2.3.5.
No, we did not exclude the first days of our simulation. We did try this but did not find a significant difference in the yearly average values when excluding the first days of the run. We added the following sentence in section 2.3.1:
Our WRF simulations were covering exactly one year, either 2002 or 2006.

Page 7, line 5: the climatological data, is it for the column or the profile? I guess it is the profile. Please provide the information.
It is for the profile. We clarified this in the paper by adding "profiles" in the text:
The CO boundary conditions of the outer domain were based on MOPITT profiles of climatological retrieved data.

Page 8, lines 26-27. There is a repetition of this information: "background simulation without emissions". Please rephrase.

We changed the text to take out the repetition and we now only describe the standard background simulation in the subsection "From model mixing ratios to emission", the other background simulation was described in the section "Sensitivity tests":

For each year also a background simulation was performed where the boundary and initial conditions are kept the same as in the simulations with emission but where emissions were switched off. The difference between these simulations represents the contribution of the emissions of Madrid to the simulated CO concentrations.

We added the following in the paragraph on sensitivity tests:

Extra background simulations were performed in order to quantify this effect: simulations with emissions outside of the 200x200 km^2 box around Madrid, and, as the normal simulation, without emissions in the urban area where the optimizations were performed.

Page 9, lines 24-27: It is not clear. Please rephrase.

We rephrased the sentences. We hope the text is clear now:

Four different filtering methods were tested to prevent outliers in the MOPITT data to influence the estimation: 1) Filtering out all MOPITT data that were more than three or 2) four standard deviations from the yearly 200x200 km^2 mean MOPITT CO concentration, or filtering out all MOPITT and WRF data at the same time and location that had a larger difference between them than 3) three (which is the default method) or 4) four standard deviations from the mean difference be- tween MOPITT and WRF at the same time and location.

Page 10: I am not sure to fully understand your discrepancy ($0.5*10^{17}$ molecules/cm^2). If I average the absolute difference between Vd-Vu from your study and Vd-Vu from P13 in 2000-2003, I find $0.45*10^{17}$ molecules/cm^2. Is it the calculation done? Please clarify this point. Same question with 0.009 and 1.04 as I do not find these values in Tab. A1.

$0.5286*10^{17}$ is the mean difference between Vd-Vu from our study with MOPITT V5 data and P13 for both 2000-2003 and 2004-2008, thus comparing each city for both time periods our study and P13 and then calculating the mean for all cities and both time periods.

0.00883 is the minimum difference we found between our results with V5 and V6 data: the difference in Vd-Vu between V5 and V6 for Sao Paulo 2000-2003.

1.014 (and not 1.04, typing error) is the maximum difference we found between our study V5 and V6: Tehran 2000-2003.

We changed the text slightly to clarify:

When the results of our approach are compared between using V5 and V6 of the data (compare Table A1 with Table A2), we find absolute discrepancies between 0.009×10^{17} and 1.014×10^{17} molecules/cm^2 with an average discrepancy of 0.3×10^{17} molec/cm^2 .

Page 10, line 20: -20%: where does the number come from?

This is a tilde: "~"20%, meant to indicate differences of around 20%. This is a rough estimate of the difference between V5 of our study and P13 in Figure 2. We changed

uncertainty to difference in the text to make this clear:
The RD estimations, however, do agree with an absolute difference of ~20% for most cities, so the method still has some value to make a rough estimation of trends in a simple and fast way.

Tab. A1 There is an error with the numbers. I think it is for example Moscow: 3.19_0.04
The "±" is missing everywhere.
Something went wrong indeed with copying the table to LaTeX. We included the ± .

Page 11, Sect 3.2.1. Did you test your results by excluding 2000 and 2001 since there is a lack of data (i.e. Jan-Feb 2000 and June-July 2001)?
No, we did not test this. As can be seen in Figure 3, left side, the variations in average total columns are indeed largest for 2000 and 2001. On the right side of the figure we show the downwind-upwind differences per year. The variation is very large, but 2000 and 2001 are not distinguishable as different from the other years. The point we want to make here is that temporal and spatial sampling differences between years can make an important difference in downwind-upwind differences. The exclusion of 2000 and 2001 would, in our opinion, not add additional information on this point.

Line 14: What does it mean? "For example, a year with below average cloud cover…"
We are not sure that we understand your question correctly here. We mean a year that is less cloudy in the summer than an average year. We hope we made it clear in the text now:
For example, a year with fewer overcast days in summer than an average year …

Page 13, line 4: "AK is scaled". It is confusing. You should specify that you are scaling an artificial AK for your test. During my first reading, I understood you wanted to artificially change the MOPITT AKs.
We changed the text now to clarify:
For Madrid, we tested this by constructing a synthetic dataset of MOPITT retrievals for the years 2000 to 2008, all based on WRF-Chem simulated CO vertical profiles over Madrid for 2002 sampled at MOPITT time and location. For each year, we constructed artificial AKs based on the MOPITT AKs. Every AK is scaled such that the annual mean sensitivity remains at the level of 2002 for each AK layer. This led to a negative difference in RD of –5% compared to the same calculation with original AKs.

Tab1 why there is only a few values underlined? Do you want to highlight something?
Indeed we wanted to highlight the method we use as standard method. We clarified this in the text:
The results of these tests are summarized in Fig. 8 and Table 1. The results of the default procedure that are shown as blue triangles in Fig. 9 are underlined in Table 1.

Page 17, line 24: 32% and in Tab A1 it is 33%
Changed, both should be 33%
Line 26 "with the increase estimated using the WRF optimization method" and in line 21, it is written -8%. Please clarify.
We found indeed a decrease using the WRF optimization method with the standard filtering. Averaged over all sensitivity tests, however, we found a positive trend. Both are stated in line 26 - 28. We changed the text about the agreement to make it more consistent:
However, when we limit this satellite-only analysis to the years 2002 and 2006,

a 5% emission increase is found ($V_d - V_u = 1.01 \times 10^{17}$ in 2002 and 1.07×10^{17} in 2006), which is in better agreement with the small increase estimated with the average of all sensitivity tests of the WRF optimization method and the relatively small decrease estimated with the standard WRF optimization method.

Fig 7. C and F are similar. Please check if the maps are correct.
The maps are correct and slightly different. The correction factors are very small, which leads to very small differences.

Page 30. What is this paragraph below figure 7?
This is part of paragraph 3.3 on emission estimation with the WRF optimization method. We moved it to the right place again.

Fig9. Write in the caption the difference between both panels.
We added this information to the caption:
upper panel: emission estimations based on EdgarV4.2 prior only; lower panel: including other prior emissions in the WRF model for optimization (see text). The uncertainty of the Edgar and MACC emission inventory estimates are estimated at 50%-200% (Kuenen et al., 2014)

Figs. A1 and Fig.A2: Add statistical values for the comparison: correlation coefficient, NMB, etc.
We added the correlation coefficient, mean absolute error and root mean square error.

Fig A1. Please improve the resolution of this figure. *done*

Page 33 and Fig. A6. Why there are less data in Figs. A6a and A6b. I think it is due to the lack of observations related to the period of the measurements. So please write the number of observations available for the comparison for each plot. What these 10000 points refer to? It is confusing. The differences between MOPITT and WRF could also be related to the difference of the initial horizontal resolution (22km _ 22km at nadir for MOPITT and 0.125_0.0625_ in the model).
As is described in Appendix B, all subplots contain the same amount of data. There are 100x100 grid cells of 2x2km^2 on which the data of MOPITT and WRF is gridded using the oversampling technique. In the shorter periods there are grid cells that contain exactly the same information as the neighboring cells, leading to more overlapping points.

Last line Appendix B. It is the same sentence in Sect 2.3.5. Do not need to repeat twice.
We deleted the double information in Appendix B.

Minor comments:
Thank you for noting, we changed our text as suggested, except if otherwise stated
Page 2, line3: quality, spatial resolution *done*
Line 6: (e.g., Beirle et al.,2011; Liu et al.,2016). Line 15: (e.g., Holloway et al., 2007; Khalil and Rasmussen,1990) Line 34: (e.g., Hooghiemstra et al.,2012a; Leeuwen van et al., 2013; Hooghiemstra et al. 2012b; Girach and Nair, 2014; Yin et al., 2015; Jiang et al., 2017) Same thing for page 4, line 8 – page 8, line 19. *done*
Page 2, line 10: at ground level at high concentration *done*
Line 16: CO is also highly dependent on seasonal variation. *This is noted in Line 15.*
Page 4, line 2: (Deeter et al., 2013; 2014) Line 3: vegetation - Deeter et al., 2009) Line 8: Deeter et al. (2014; 2016). *done*

Page 7, line 16: we used emissions from the EdgarV4.2 *done*
Page 8, line 21: "coarser spatial resolutions": Please provide these resolutions. *done: 0.1x0.1 degree*
Page 10, line 34: weighting *done*
Page 11, line 20: need to correct the numbers: 10ˆ16 10ˆ17 *done*
Page 12, line 24: (from surface to 800 hPa). *done*
Page 12, line 25 & Fig. A4: AK area. Do you mean AK vector? *No we did not mean AK vector. We described the AK area, as is done first by Rodgers (2000), after the colon in line 25.*
Page 15: problem in inversion studies (Jacob et al., 2016). *done*
Page 20 line 8: Do not begin the sentence with "Or," *done*
Fig1. Please add the location of Madrid on the map. *done*
Figs. 2 & 5. It is very nice and interesting. *Thank you*
figA6. Add labels (a), (b), (c) and (d) on the scatterplots. *These labels are already included in the lower left corner*
Tab. A1 & A2. Write: "… from this study and Pommier et al. (2013). The values from Pommier et al. (2013) are provided in parenthesis". *done*